Research on Improved Dual Channel Medical Short Text Intention Recognition Algorithm

Chao Wang
Xi'an Technological University
College of Computer Science and Technology
Xi'an, China
E-mail:1501873640@qq.com

Fei Xu
Xi'an Technological University
College of Computer Science and Technology
Xi'an, China
E-mail:xufei@xatu.edu

Yongyong Sun
Xi'an Technological University
College of Computer Science and Technology
Xi'an, China
E-mail:yongsunjd@126.com

Abstract—**The increasing application of medical robots in the healthcare sector underscores the critical importance of intent recognition in enhancing the interaction and assistance capabilities of these robots. Traditional intent recognition methods utilize convolutional neural networks (CNNs) for text analysis but often fall short in capturing global features, resulting in incomplete information. To address this challenge, this paper introduces an innovative approach by combining an enhanced CNN with bidirectional gated recurrent units (BiGRU) to construct a dual-channel short-text intent recognition model. This model effectively leverages both local and global features to more accurately comprehend user needs and intentions. Experimental results demonstrate that this model excels, achieving an accuracy rate of 96.68% and an F1 score of 96.67% on the THUCNews_Title dataset. In comparison to conventional intent recognition models, it exhibits significantly improved performance, thereby providing substantial support for medical robots in patient care and assisting healthcare professionals.**

Keywords-Intention Recognition; Albert; Bigru; Dual Channel

I. INTRODUCTION

Natural language understanding(NLU) plays a fundamental role in robot question-answering systems in the medical field. Exceptional intent recognition modules help simplify the complexity of NLU, allowing robots to more effectively process text by categorizing intricate questions into the relevant intents. Medical question-answering is a focal point in robotics research within the medical domain due to the highly specialized nature of medical knowledge. Accurately identifying the intent of questions enables robots to better integrate medical knowledge, thus enhancing search result performance. In comparison to systems without intent recognition or those with suboptimal performance, outstanding medical question-answering modules can significantly alleviate the workload of subsequent robot tasks. From both input and output perspectives, intent recognition tasks can be regarded as text classification tasks within the realm of natural language processing, providing foundational support for robot work in the medical domain.

With the continuous advancement of neural network technologies, researchers have dedicated substantial efforts to improve intent recognition, particularly in the context of medical question-answering. Short-text medical queries, characterized by their brevity and concise information, often pose challenges for traditional intent recognition methods. The accuracy of natural language understanding is paramount to the performance of the entire question-answering system. Inaccurate comprehension in the early

DOI: 10.2478/ijanmc-2023-0079

stages of information retrieval can lead to subsequent inaccurate responses. Therefore, the aim of this study is to provide more accurate natural language understanding tools, especially for handling interrogative sentences in the domain of medical robotics. In this paper, we introduce a dual-channel medical short-text intent recognition model, denoted as the AB-CNN-BGRU-att model. This model combines TextCNN with BiGRU-att and employs multiple pooling strategies. The BiGRU-att module employs a dual-channel approach to capture features at different levels, thereby capturing the global information within the text. Simultaneously, TextCNN leverages different-sized convolution kernels and pooling strategies to extract a broader array of local features. Experimental results demonstrate that the AB-CNN-BGRU-att model outperforms other popular intent recognition models, particularly in the context of medical robotics applications. This model significantly enhances a robot's ability to comprehend interrogative sentences.

II. RELATED WORKS

The document [1] initially introduced Convolutional Neural Networks (CNN), originally used in computer vision, as the TextCNN model, a classic model in text classification. Later, WANG Haitao et al. [2] addressed TextCNN's shortcomings in handling short texts by employing non-linear sliding methods and N-gram models. Ma Sidan et al. [3] improved Word2vec by utilizing text similarity for classification. Subsequently, Sun Hong et al. [4] and Chi Haiyang et al. [5] used BERT as an embedding layer, incorporating BiGRU to capture global sentence features, and utilized attention mechanisms for classification, demonstrating improvements on their respective datasets. Due to the large size of the BERT model, Wen Chaodong et al. [6] and Zeng Cheng [7] proposed using the ALBERT model as an enhancement. Wen Chaodong's experiments with the ALBERT-BiGRU model exceeded the performance of Word2vec and GloVe. Zeng Cheng, using the ALBERT model, demonstrated improved F1 values compared to other models by connecting the CNN layer and feeding it into the BiGRU layer for classification.

Recent scholars [8-10] emphasize the significance of both local and global features in short text corpora. Their multi-channel approach processes inputs independently from the embedding layer and merges features for classification. Additionally, Wu Di et al. [11] proposed an enhanced embedding layer in a dual-channel model by combining static and dynamic word vectors from ELMo and GloVe, outperforming traditional models on datasets such as IMDB.

III. MODEL FRAMEWORK

The TextCNN model uses CNN for text feature extraction but overlooks the entire sentence context. To address this, we augment BERT TextCNN with a BiGRU network to capture global features. We replace the heavier BERT with the lighter ALBERT, which still generates rich word vectors but with fewer parameters. These vectors feed into TextCNN and BiGRU to respectively extract local and global features. TextCNN utilizes various kernel sizes and pooling strategies, while BiGRU captures global features. These features are merged, and using Dropout and softmax, probability values are computed for multi-classification. Refer to Figure 1 for the model architecture.

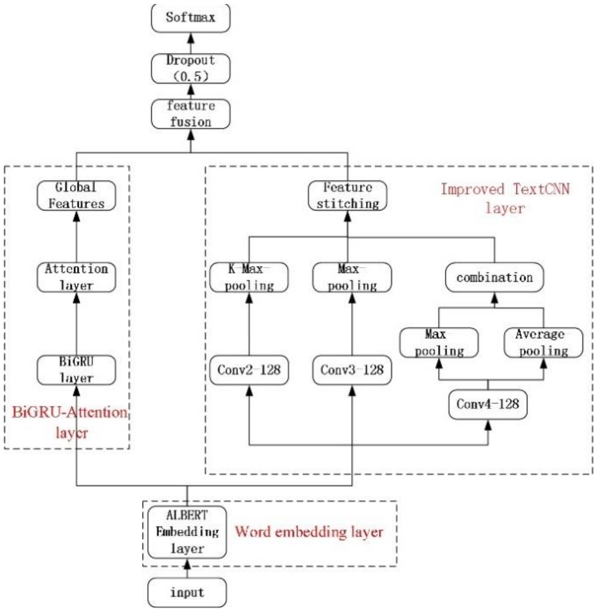

Figure 1. Architecture diagram of AB-CNN-BGRU-att model

A. Word embedding layer

Several Word embedding models are widely used, including Word2vec, GloVe, and BERT. Among these, BERT has gained recognition in numerous experiments within the NLP community, being considered one of the top models. ALBERT, a variation of BERT, simplifies the original BERT while maintaining similar performance. Official data from the ALBERT paper reveals that it achieves comparable performance to BERT base across several representative tasks, yet with significantly fewer parameters—six times fewer—and nearly three times faster processing time. Consequently, this paper adopts ALBERT for the Word embedding layer due to its efficiency, as illustrated in Figure 2.

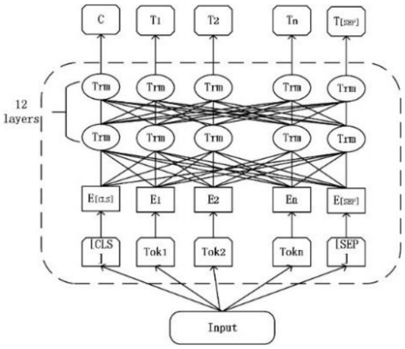

Figure 2. ALBERT model structure

When text enters the Word embedding layer, it's marked with [CLS] and [SEP] to show sentence boundaries. The resulting serialized text generates the En vector, which, processed by the Transformer encoder, produces Tn from its features. ALBERT and BERT both employ the Transformer's encoder section, composed of multiple identical network layers featuring residual connections between the "Multi Head Attention" and "FeedForward" layers. The "Multi Head Attention" layer functions on input vectors Q, K, and V, derived from text queries, keys, and values in the sequence, with equations (1) to (3) defining the specific computations.

$$head_t = Attention(QW_t^Q, KW_t^K, VW_t^V), t \in (1,2,...,h), \quad (1)$$

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{(d_t)}})V \quad (2)$$

Merge the resulting matrices:

$$MultiHead(Q,K,V) = Concat(head_1, head_2, ..., head_h)W^0 \quad (3)$$

W^0 represents the weight matrix to ensure the final matrix's dimensions align with the sequence length, W_t^Q , W_t^K , W_t^V represents the weight matrices for individual Q, K, and V vectors, while d_t denotes their dimensional size.

B. BiGRU-att Module

To capture global text features and enhance the model's grasp of the text's core concept, this study integrates a BiGRU layer following ALBERT's word vector output. This layer extracts comprehensive feature details for the entire sentence. The network structure of the GRU is depicted in Figure 3.

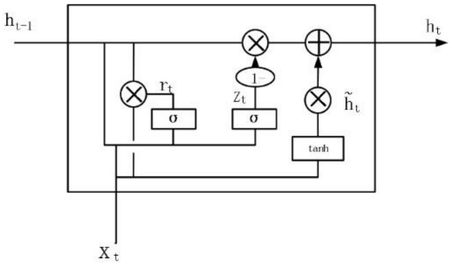

Figure 3. GRU network structure

The standard GRU's hidden state is unidirectional, focusing solely on the present input state without considering the impact of the text context on this state. This unidirectional nature fails to capture how subsequent information affects preceding states. To address this limitation, this paper employs BiGRU, a variant of GRU. BiGRU integrates two GRU layers with opposite directions, allowing output information to be influenced by both directional outcomes. Formula (4) demonstrates the final output result, while Figure 4 illustrates the BiGRU model structure.

$$h_t^{(i)} = [\vec{h}_t^{(i)}, \overleftarrow{h}_t^{(i)}] \quad (4)$$

In the above equation, $\vec{h}_t^{(i)}$ represents the information obtained by the i-th text passing through the forward GRU, and $\overleftarrow{h}_t^{(i)}$ represents the information obtained by the i-th text passing through the backward GRU. $h_t^{(i)}$ is the final result obtained from this text through BiGRU.

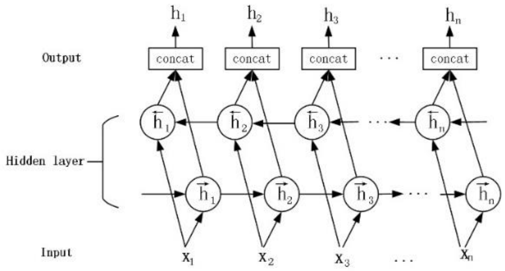

Figure 4. BiGRU network structure

The attention mechanism assigns weights to words in the text to prioritize crucial features, enabling the model to concentrate on words with higher weight scores and enhance classification accuracy. In this process, the Attention layer computes word weights for each BiGRU output vector, generating a final sentence representation by the weighted sum of these scores and corresponding position feature vectors. This BiGRU Attention layer enables the model to autonomously emphasize significant words with higher weight scores, thereby improving its ability to capture global features in the input text.

C. Improved TextCNN Module

The revised TextCNN model includes multiple convolutional layers with varied sizes, diverse pooling layers, and fully connected layers, an advancement from the original TextCNN model. The model architecture is depicted in Figure 5.

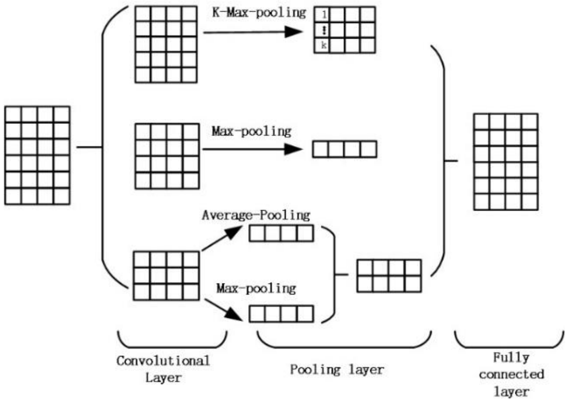

Figure 5. Improved TextCNN model structure

The enhanced TextCNN layer conducts convolution by examining text features using various convolution kernel sizes, aligned with the Token Embedding's dimension, and the row information of the vector matrix represents words. When the width of the convolution kernel is inconsistent with the dimension of Token Embedding, the convolution kernel cannot extract complete word information. This matrix acquires the characteristic mapping matrix. The enhanced TextCNN layer conducts convolutions on text features using diverse kernel sizes. The kernel width matches the Token Embedding dimension, with rows in the vector matrix representing words. If the kernel width mismatches the Token Embedding dimension, complete word information might not be extracted. Through a nonlinear Activation function, this matrix obtains the feature mapping matrix $c = [c_1, c_2, \cdots, c_n]$.The characteristic formula is presented in Formula (5).

$$c_i = f(w \cdot x_{i:i+h-1} + b) \tag{5}$$

Where, f is the Activation function, w is the weight matrix of the convolution kernel, and b is the offset term.

In short text questions, the TextCNN model focuses on extracting local features due to its dual-channel structure and limited sentence length. Unlike common choices such as kernel sizes 3, 4, and 5, this model selects sizes of 2, 3, and 4 for the convolutional kernels. For kernel size 2 convolution, the model uses K-Max pooling. Which selects the top K scores during pooling, capturing more abundant information compared to the typical maximum pooling method. The latter overlooks repeating features, seeing them only once, while K-Max pooling retains relative order information between some features by retaining K higher-scored features.

For kernel size 3, the model uses max pooling, focusing on essential text features by discarding weaker ones, minimizing noise, and emphasizing keywords.

Using a kernel size of 4, the model employs both maximum and average pooling strategies. Concatenating the resulting features is beneficial as maximum pooling focuses on the highest-scored feature, whereas average pooling considers each word's information.

After utilizing different kernel sizes and corresponding pooling operations, diverse local features are acquired. To prevent overfitting, a dropout layer follows the TextCNN pooling layer, enhancing the model's generalization. These features, combined with global features from the BiGRU-att module, form the final feature vector. Classification results are determined through the final fully connected layer, as outlined in Formula (6).

$$Z = soft\max(W_Z \cdot F + b) \tag{6}$$

Among these, Z stands for the predicted intention tag result, $soft\max$ represents the Activation function, W_Z indicates the weight of the fully connected layer, F is the final feature vector, and b represents the offset term.

D. AB-CNN-BGRU-att algorithm

The AB-CNN-BGRU-att (ALBERT-TextCNN-BiGRU-attention) algorithm determines intention labels' probabilities for intention recognition by analyzing the input text corpus. Its detailed process is depicted in Algorithm 1.

ALGORITHM I. AB-CNN-BGRU-ATT ALGORITHM FLOW

```
Algorithm: AB-CNN-BGRU-att algorithm
Input: $S = (s_1, s_2, s_3, \cdots s_n)$, $s$ is the input text sequence
Output: Intention identification label results
1. Data preprocessing, importing training sets, testing sets
2. Load the ALBERT model to obtain dynamic word vectors Token
3. $conv_{output_{1\sim n}} = Conv_{1\sim n}(T);$
4. $pooling_{output_{1\sim n}} = Pooling(Conv_{output_{1\sim n}});$
5. $cnn\_output = Concat(pooling_{output_{1\sim n}});$
6. $forward = GRU(T);$
7. $backward = GRU(T);$
8. $bigru\_output = Concat(forward, backward);$
9. $output = Concat(cnn\_output, bigru\_output);$
10. $dropout = Dropout(output);$
11. $dense = Dense(dropout);$
12. $out = Softmax(dense);$
13. $END$.
```

IV. EXPERIMENT AND RESULT ANALYSIS

A. Experimental data

The paper uses two datasets for experiments. The first one, THUCNews_Title, is drawn from THUCNews, containing 200,000 titles, each not exceeding 30 characters. It covers 10 categories. Table 1 illustrates the THUCNews_Title dataset.

The study focuses on common medical conditions. The KUAKE-QIC dataset, sourced from Alibaba Tianchi Laboratory, validates the model's performance. This dataset aims to improve search result relevance in medical queries, crucial in a field with specialized knowledge. It includes 11 categories. There are 6931 training, 1955 validation, and 1994 test samples. Approximately 96% of the data (6684 samples) contain less than 30 words, fitting the experimental criteria for short medical text datasets. Table 1 displays the KUAKE-QIC dataset.

TABLE I. EXPERIMENTAL DATASET

Name	Training Set	Test Set	Validation Set	Category	Total
KUAKE-QIC	6931	1994	1955	11	10880
THUCNews_Title	180000	10000	10000	10	200000

B. Parameter settings

The key parameters of the improved version of TextCNN in the AB-CNN-BGRU att model are as follows: word vector dimension is 384, activation function uses ReLu, learning rate is 1e-5, Dropout is 0.5, and batch size is 128. The key parameters of

BiGRU att in the AB-CNN-BGRU att model are as follows: hidden layer size is 256, word vector dimension is 384, activation function uses ReLu, Dropout is 0.2, and batch size is 128.

C. Experimental result

By tuning the model's hyperparameters and training it on the THUCNews_Title dataset, the model was tested on the THUCNews_Title test set. The obtained results for each category were compared. It's observed that the model achieves classification scores above 90% for each category. Notably, technology-related texts pose higher complexity and uncommon vocabulary, while stocks and social texts share similarities with other labels, leading to potential confusion and reduced accuracy. Overall, the model demonstrates its ability to accurately identify intentions in short texts, effectively interpreting text intentions despite limited sentence information and a concise corpus. Figure 6 displays the validation results of the model.

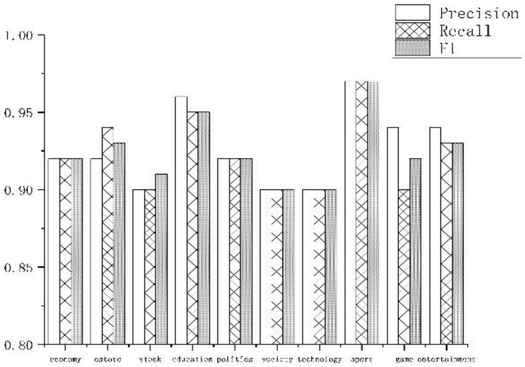

Figure 6. Model validation results

In the KUAKE-QIC dataset experiment, the model was trained for 20 epochs. The test set achieved an accuracy of 86.02, positioning it as the third-ranked model in the CBLUE3.0 ranking. The top-ranking model achieves an accuracy of 87.0117, followed by the second and third models with accuracies of 86.0589 and 85.9084, respectively.

D. Comparative experiment

In an experiment comparing the BiGRU att and BiLSTM att layers, this study maintained consistent parameters across two network layers. The comparison encompassed factors like average time per epoch, total training time, final accuracy, and F1 value. Figure 7 displays the contrast in epoch times, while Table 3 offers a comprehensive overview. Notably, the BiGRU att layer significantly outperformed the BiLSTM att layer in training efficiency. Although both layers achieved comparable accuracy and F1 scores, the study opts for the BiGRU-att layer due to its superior training efficiency and outcomes.

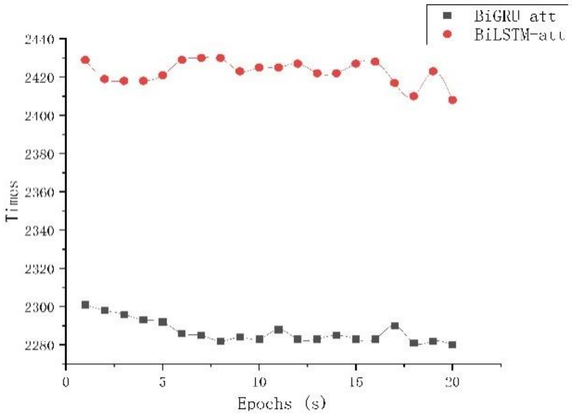

Figure 7. Comparison of Network Time

TABLE II. COMPARISON BETWEEN BIGRU-ATT AND BILSTM-ATT

Network Layer	Average Duration	Total Duration	Acc%	F1%
BiGRU-att	2286.9s	45738s	90.83	90.64
BiLSTM-att	2422.55s	48451s	90.45	90.41

To confirm our model's superiority under identical conditions to other models, we conducted comparative experiments using the THUCNews_Title dataset. Details of the models used in the experiments are outlined below:

SAttBiGRU: Utilizes BiGRU to capture global features and enhances text features by applying Self Attention, providing richer feature information for classification.

Self-Attention-CNN: Combines Self Attention with the fundamental TextCNN. It applies weighting using Self Attention to compact text information from the TextCNN's embedding layer. Following max pooling, the fully connected layer outputs classification results.

BiGRU-MCNN: Global features are extracted via BiGRU, while detailed local features are obtained through multi-channel CNN. The model then merges these two feature types and utilizes a fully connected layer to generate classification outcomes.

MC-AttCNN-AttBiGRU: Initially employs the attention mechanism to weigh multi-channel CNN and BiGRU. Subsequently, it concatenates the derived text feature vectors and conducts classification via a fully connected layer.

TABLE III. COMPARISON OF EXPERIMENTAL RESULTS

Model	Acc%	Pre%	Recall%	F1%
SAttBiGRU	96.16	96.20	96.16	96.17
Self-Attention-CNN	94.85	94.89	94.85	94.85
BiGRU-MCNN	95.43	95.45	95.43	95.43
MC-AttCNN-AttBiGRU	95.93	95.98	95.93	95.93

The results in Table 4 demonstrate the performance superiority of the AB-CNN-BGRU-att model over other models. Across various indicators using the THUCNews_Title dataset, this model exhibits a consistent improvement of one to two percentage points compared to the best-performing existing models. These comparative findings substantiate the advantageous performance of the AB-CNN-BGRU-att model proposed in this study.

E. Ablation experiment

Ablation experiments were performed to assess the efficiency of the proposed model for short text classification. Each local network element - TextCNN, enhanced TextCNN, BiGRU att, and AB-CNN-BGRU-att - underwent individual analysis in these.ALBERT was utilized as the Word embedding layer. The findings from the ablation experiment are summarized in Table 9:

TABLE IV. RESULTS OF ABLATION EXPERIMENT

Model	Acc%	Pre%	Recall%	F1%
TextCNN	89.96	89.90	89.96	89.90
Improved TextCNN	94.85	94.89	94.85	94.85
BiGRU-att	94.00	94.17	94.00	94.90
AB-CNN-BGRU-att	96.68	96.68	96.67	96.67

Table 5 demonstrates that the basic TextCNN model yielded unsatisfactory classification results with all indicators below 90%. This poor performance might stem from TextCNN's inefficiency in handling short texts. In contrast, the improved TextCNN model significantly enhanced all indicators. Employing various convolution kernel sizes and pooling strategies proved effective in obtaining richer local features, enhancing the model's performance. However, the BiGRU-att model exhibited slightly lower performance compared to the improved TextCNN model, highlighting the importance of global features in recognizing intentions, the AB-CNN-BGRU-att model, integrating local and global features, demonstrated an enhancement of almost two percentage points over the improved TextCNN and BiGRU-att models.

V. CONCLUSIONS

This paper introduces a dual-channel intent recognition model for medical short texts by combining Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit with Attention (BiGRU-Att). It also incorporates ALBERT, BiGRU attention, and an enhanced TextCNN model. The model processes vectors obtained from ALBERT separately, sending them to the BiGRU-Att network model for global feature extraction and the TextCNN model for local feature extraction using multiple pooling strategies and a hybrid pooling approach. After merging these two types of features, a classification result is obtained through a fully connected layer with softmax activation. This model's performance is evaluated against four other models using publicly accessible datasets.

Comparative experimental data clearly demonstrate the superior performance of the proposed model across various evaluation metrics. The experimental data also demonstrate the model's capacity to yield more precise intent recognition outcomes, which are crucial for the tasks performed by medical robots in the healthcare domain.

While the model in this study excels in short-text intent recognition for medical domain robots, it still heavily relies on extensive annotated datasets as the foundation. Subsequent work will explore semi supervision learning approaches to reduce manual annotation efforts and simultaneously enhance model performance, thus better supporting the applications of medical domain robots.

ACKNOWLEDGMENT

One of us (R.B.G.) thanks New Network and Detection Control National Joint Local Engineering Laboratory Fund Project.

REFERENCES

[1] Yoon Kim. Convolutional Neural Networks for Sentence Classification. [J]. CoRR, 2014 ,abs/1408. 5882(abs/ 14 08.5882).

[2] WANG Haitao, HE Jie, ZHANG Xiaohong, LIU Shufen. A Short Text Classification Method Based on N-Gram and CNN [J]. Chinese Journal of Electronics, 2020,(02): 248-254.

[3] MA Si-dan，LIU Dong-su. Text Classification Method Based on Weighted Word2vec [J]. Information Science, 2019, (11):38-42.

[4] SUN Hong, CHEN Qiang-yue. Chinese Text Classification Based on BE Ｒ T and Attention[J]. Journal of Chinese Computer Systems, 2022, (01):22-26.

[5] CHI Haiyang, YAN Xin, ZHOU Feng, XU Guangyi, ZHANG Lei. An online health community user intention identification method based on BERT-BiGRU-Attention [J]. Journal of Hebei University of Science and Technology, 2020, (03):225-232.

[6] WEN Chaodong, ZENG Cheng, REN Junwei , ZHANG Yan. Patent text classification based on ALBERT and bidirectional gated recurrent unit [J]. Journal of Computer Applications, 2021, (02):407-412.

[7] ZENG Cheng, WENE Chaodong, SUN Yumin, PAN Lie, HE Peng. Motional Analysis of Bullet Screen Text Based on ALBERT-CRNN [J]. Journal of Zhengzhou University(Natural Science Edition), 2021, 53(3): 1-8.

[8] LI Yang, DONG Hongbin. Text sentiment analysis based on feature fusion of convolution neural network and bidirectional long short-term memory network [J]. Journal of Computer Applications, 2018, (11):3075-3080.

[9] LI Qihang, LIAO Wei, MENG Jingwen. Dual-Channel DAC-RNN Text Classification Model Based on Attention Mechanism [J/OL]. Computer Engineering and Applications: 1-9(2021-04-21) [2022-01-25].

[10] SONG Zhongshan,NIU Yue,ZHENG Lu,TIE Jun, JIANG Hai. Multiscale double-layer convolution and global feature text classification model [J]. Computer Engineering and Applications:1-11.

[11] WU Di, WANG Ziyu, ZHAO Weichao. ELMo-CNN-BiGRU Dual-Channel Text Sentiment Classification Model [J]. Computer Engineering, 2022, (08):105-112.